GenCore version 4.5

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:55 ; Search time 91.75 Seconds
(without alignments)
14.079 Million cell updates/sec

Title: US-09-331-631A-8_COPY_80_119
Perfect score: 225
Sequence: 1 PEDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	221	98.2	588	1	VCLB_GOSHI	P09801 gossypium h
2	142	63.1	605	1	VCLA_GOSHI	P09799 gossypium h
3	86	38.2	1905	1	TAGB_DICDI	P54683 dictyosteli
4	83	36.9	429	1	APA4_MACFA	P33621 macaca fasc
5	83	36.9	1403	1	PRO_DROME	P29617 drosophila
6	82	36.4	339	1	TF2D_HUMAN	P20226 homo sapien
7	82	36.4	401	1	APA4_PAPAN	Q28758 papio anubi
8	78	34.7	648	1	KAPC_DICDI	P34099 dictyosteli
9	77.5	34.4	2124	1	Y192_HUMAN	Q93074 homo sapien
10	77	34.2	758	1	YM38_YEAST	Q03825 saccharomyc
11	76.5	34.0	467	1	INVO_MOUSE	P48997 mus musculu
12	76.5	34.0	738	1	YKF4_YEAST	P35732 saccharomyc
13	76	33.8	905	1	SNF5_YEAST	P18480 saccharomyc
14	75	33.3	1023	1	CLOC_DROME	O61735 drosophila
15	75	33.3	1154	1	WC1_NEUCR	Q01371 neurospora
16	74	32.9	550	1	CCF_DROME	P41046 drosophila
17	73	32.4	1505	1	SIMA_DROME	Q24167 drosophila
18	72.5	32.2	262	1	GDA1_WHEAT	P04721 triticum ae
19	72.5	32.2	307	1	GDA9_WHEAT	P18573 triticum ae
20	72	32.0	538	1	CK11_YEAST	P23291 saccharomyc
21	71	31.6	286	1	GDA0_WHEAT	P02863 triticum ae
22	71	31.6	1365	1	SUZ2_DROME	P25172 drosophila
23	70.5	31.3	186	1	GDA8_WHEAT	P04728 triticum ae
24	70.5	31.3	255	1	LP61_EIMTE	P15714 eimeria ten
25	70	31.1	305	1	HOG1_HORVU	P17990 hordeum vul
26	70	31.1	313	1	GDA7_WHEAT	P04727 triticum ae
27	70	31.1	319	1	GDA5_WHEAT	P04725 triticum ae
28	70	31.1	3828	1	TRX_DROVI	Q24742 drosophila
29	69	30.7	47	1	AGRP_LUFCY	P56568 luffa cylin
30	69	30.7	296	1	GDA6_WHEAT	P04726 triticum ae
31	69	30.7	924	1	CIKB_DROME	P17970 drosophila
32	69	30.7	1090	1	NIT4_NEUCR	P28349 neurospora
33	68	30.2	398	1	PF21_ARATH	Q04088 arabidopsis
34	67.5	30.0	2703	1	NOTC_DROME	P07207 drosophila
35	67	29.8	966	1	SSN6_YEAST	P14922 saccharomyc
36	67	29.8	1319	1	MN1_HUMAN	Q10571 homo sapien
37	67	29.8	1898	1	TRHY_HUMAN	Q07283 homo sapien
38	67	29.8	2038	1	FSH_DROME	P13709 drosophila
39	66	29.3	291	1	GDA2_WHEAT	P04722 triticum ae
40	66	29.3	297	1	GDA4_WHEAT	P04724 triticum ae
41	66	29.3	468	1	YQC1_CAEEL	Q09260 caenorhabdi
42	66	29.3	623	1	DSH_DROME	P51140 drosophila
43	66	29.3	1023	1	GLT_DROME	P33438 drosophila
44	66	29.3	1073	1	HR38_DROME	P49869 drosophila
45	66	29.3	1329	1	FTSK_ECOLI	P46889 escherichia

ALIGNMENTS

```
RESULT   1
VCLB_GOSHI
ID   VCLB_GOSHI     STANDARD;      PRT;   588 AA.
AC   P09801;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Malvales; Malvaceae; Gossypium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;
RT   "Developmental biochemistry of cottonseed embryogenesis and
RT   germination. XVIII. cDNA and amino acid sequences of the members of
RT   the storage protein families.";
RL   Plant Mol. Biol. 7:475-489(1986).
CC   -!- FUNCTION: SEED STORAGE PROTEIN.
CC   -!- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC       BODIES.
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16891; AAA33071.1; -.
DR   PIR; A30838; FWCNAB.
DR   HSSP; P50477; 1CAX.
DR   INTERPRO; IPR001113; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal.
FT   SIGNAL       1     25
FT   CHAIN       26    588       VICILIN C72.
SQ   SEQUENCE   588 AA;  69729 MW;  63E699B29AB8ADEB CRC64;


  Query Match             98.2%;  Score 221;  DB 1;  Length 588;
  Best Local Similarity   97.5%;  Pred. No. 1.3e-15;
  Matches   39;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 PEDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
           |||||||||||||||||||||:|||||||||||||||||
Db      80 PEDPQRRYEECQQECRQQEERQRPQCQQRCLKRFEQEQQQ 119


RESULT   2
VCLA_GOSHI
ID   VCLA_GOSHI     STANDARD;      PRT;   605 AA.
```

```
AC   P09799;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Malvales; Malvaceae; Gossypium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT   "Developmental biochemistry of cottonseed embryogenesis and
RT   germination. XIX. Sequences and genomic organization of the alpha
RT   globulin (vicilin) genes of cottonseed.";
RL   Plant Mol. Biol. 9:533-546(1987).
CC   -!- FUNCTION: SEED STORAGE PROTEIN.
CC   -!- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC       BODIES.
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19378; AAA33069.1; -.
DR   PIR; S06398; S06398.
DR   HSSP; P50477; 1CAX.
DR   INTERPRO; IPR001113; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal.
FT   SIGNAL        1     23
FT   CHAIN        24    605       VICILIN GC72-A.
SQ   SEQUENCE   605 AA;  71049 MW;  C9DB9371C976953B CRC64;


  Query Match             63.1%;  Score 142;  DB 1;  Length 605;
  Best Local Similarity   53.8%;  Pred. No. 1.1e-07;
  Matches   21;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 EDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
           |||||||::|:| |:|:| | :| |:| | :::|::|||
Db      78 EDPQRRYQDCRQHCQQEERRLRPHCEQSCREQYEKQQQQ 116


RESULT   3
TAGB_DICDI
ID   TAGB_DICDI     STANDARD;      PRT;  1905 AA.
AC   P54683;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PRESTALK-SPECIFIC PROTEIN TAGB PRECURSOR (EC 3.4.21.-).
GN   TAGB.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Dictyosteliida; Dictyostelium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=95262903; PubMed=7744252;
RA   Shaulsky G., Kuspa A., Loomis W.F.;
RT   "A multidrug resistance transporter/serine protease gene is required
RT   for prestalk specialization in Dictyostelium.";
RL   Genes Dev. 9:1111-1122(1995).
CC   -!- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE
CC       INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY
CC       S8; ALSO KNOWN AS THE SUBTILASE FAMILY.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
CC       TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
CC   -!- SIMILARITY: STRONG, TO TAGC.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U20432; AAA62212.1; -.
DR   HSSP; P13569; 1NBD.
DR   DICTYDB; DD02059; TAGB.
DR   INTERPRO; IPR000209; -.
DR   INTERPRO; IPR001140; -.
DR   INTERPRO; IPR001617; -.
DR   PFAM; PF00664; ABC_membrane; 1.
DR   PFAM; PF00005; ABC_tran; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
KW   Signal.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?   1905       PRESTALK-SPECIFIC PROTEIN TAGB.
FT   DOMAIN      378    700       PROTEASE.
FT   DOMAIN        ?   1905       ABC TRANSPORTER.
FT   TRANSMEM   1011   1031       POTENTIAL.
FT   TRANSMEM   1076   1096       POTENTIAL.
FT   TRANSMEM   1121   1141       POTENTIAL.
FT   TRANSMEM   1210   1230       POTENTIAL.
FT   TRANSMEM   1309   1329       POTENTIAL.
FT   TRANSMEM   1332   1352       POTENTIAL.
FT   ACT_SITE    387    387       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    432    432       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    695    695       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1553   1560       ATP (POTENTIAL).
FT   DOMAIN       63     67       POLY-GLN.
FT   DOMAIN       95    104       POLY-ASN.
FT   DOMAIN      107    134       POLY-ASN.
FT   DOMAIN      311    321       POLY-SER.
FT   DOMAIN      833    839       POLY-SER.
FT   DOMAIN      838    844       POLY-GLY.
FT   DOMAIN      871    876       POLY-LEU.
FT   DOMAIN     1012   1015       POLY-ILE.
FT   DOMAIN     1386   1389       POLY-GLU.
FT   DOMAIN     1398   1404       POLY-GLY.
FT   DOMAIN     1445   1450       POLY-ASN.
FT   DOMAIN     1765   1779       POLY-ASN.
FT   DOMAIN     1782   1785       POLY-SER.
FT   DOMAIN     1807   1812       POLY-PRO.
FT   DOMAIN     1815   1860       POLY-GLN.
FT   DOMAIN     1872   1878       POLY-PRO.
FT   CARBOHYD    594    594       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    621    621       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    672    672       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    747    747       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    823    823       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1172   1172       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1522   1522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1658   1658       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1905 AA;  212518 MW;  B8E223FA8B9AE13C CRC64;


  Query Match             38.2%;  Score 86;  DB 1;  Length 1905;
  Best Local Similarity   42.5%;  Pred. No. 0.12;
  Matches   17;  Conservative   12;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 PEDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
```

```
          |  ||  : |: :|: :||:|:|| | ||:  :: :|:|||
Db   1809 PPPPQEQQEQQEQQQQQQQEQQQQQQQQQQQQQQQQQQQ 1848


RESULT   4
APA4_MACFA
ID   APA4_MACFA     STANDARD;      PRT;   429 AA.
AC   P33621;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   APOLIPOPROTEIN A-IV PRECURSOR (APO-AIV).
GN   APOA4.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEUKOCYTE;
RX   MEDLINE=93192330; PubMed=8448212;
RA   Osada J., Pocovi M., Nicolosi R.J., Schaefer E.J., Ordovas J.M.;
RT   "Nucleotide sequences of the Macaca fascicularis apolipoprotein C-III
RT   and A-IV genes.";
RL   Biochim. Biophys. Acta 1172:335-339(1993).
CC   -!- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND
CC       CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN
CC       LIPASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR
CC       COMPONENT OF HDL AND CHYLOMICRONS.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: SECRETED IN PLASMA.
CC   -!- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH
CC       22-MER IS ACTUALLY A TANDEM ARRAY OF TWO, A AND B, RELATED 11-
CC       MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-
CC       HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY
CC       THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL
CC       ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.
CC   -!- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X68361; CAA48421.1; -.
DR   PIR; S29565; S29565.
DR   PIR; S30195; S30195.
DR   HSSP; P02649; 1NFO.
DR   INTERPRO; IPR000074; -.
DR   PFAM; PF01442; Apolipoprotein; 1.
KW   Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal.
FT   SIGNAL        1     20       BY SIMILARITY.
FT   CHAIN        21    429       APOLIPOPROTEIN A-IV.
FT   DOMAIN       33    330       13 X 22 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT       33     54       1.
FT   REPEAT       60     81       2.
FT   REPEAT       82    103       3.
FT   REPEAT      115    136       4.
FT   REPEAT      137    158       5.
FT   REPEAT      159    180       6.
FT   REPEAT      181    202       7.
FT   REPEAT      203    224       8.
FT   REPEAT      225    246       9.
FT   REPEAT      247    268       10.
FT   REPEAT      269    286       11.
FT   REPEAT      287    308       12.
FT   REPEAT      309    330       13.
FT   DOMAIN      372    420       GLU/GLN-RICH.
SQ   SEQUENCE   429 AA;  49876 MW;  3D458F551D0DB60C CRC64;

  Query Match             36.9%;  Score 83;  DB 1;  Length 429;
  Best Local Similarity   46.2%;  Pred. No. 0.066;
  Matches   18;  Conservative   10;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 EDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
           |  | : :| :|| :|||:|:|| | |||  :: ||:|:|
Db     378 EQQQEQQQEQEQEQQQQQEQQQQQEQQREQQQQEQQQEQ 416


RESULT   5
PRO_DROME
ID   PRO_DROME      STANDARD;      PRT;  1403 AA.
AC   P29617;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROTEIN PROSPERO.
GN   PROS.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92069760; PubMed=1720353;
RA   Vaessin H., Grell E., Wolff E., Bier E., Jan L.Y., Jan Y.N.;
RT   "Prospero is expressed in neuronal precursors and encodes a nuclear
RT   protein that is involved in the control of axonal outgrowth in
RT   Drosophila.";
RL   Cell 67:941-953(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92171948; PubMed=1540176;
RA   Matsuzaki F., Koizumi K., Hama C., Yoshioka T., Nabeshima Y.;
RT   "Cloning of the Drosophila prospero gene and its expression in
RT   ganglion mother cells.";
RL   Biochem. Biophys. Res. Commun. 182:1326-1332(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93083413; PubMed=1842358;
RA   Chu-Lagraff Q., Wright D.M., McNeil L.K., Doe C.Q.;
RT   "The prospero gene encodes a divergent homeodomain protein that
RT   controls neuronal identity in Drosophila.";
RL   Development Suppl. 2:79-85(1991).
RN   [4]
RP   SIMILARITY TO C.ELEGANS CEH-26.
RX   MEDLINE=94212446; PubMed=7909177;
RA   Buerglin T.R.;
RT   "A Caenorhabditis elegans prospero homologue defines a novel domain.";
RL   Trends Biochem. Sci. 19:70-71(1994).
CC   -!- FUNCTION: INVOLVED IN THE CONTROL OF OTHER NEURONAL PRECURSOR
CC       GENES AS WELL AS AXONAL OUTGROWTH AND PATHFINDING OF NUMEROUS
CC       CENTRAL AND PERIPHERAL NEURONS. IT IS PROBABLY GENERALLY REQUIRED
CC       FOR PROPER NEURONAL DIFFERENTIATION OF MOST OR ALL NEURONS & THEIR
CC       PRECURSORS IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS, BUT NOT
CC       FOR THE SPECIFICATION OF IDENTITY. PROSPERO PROTEIN MAY REGULATE
CC       TRANSCRIPTION BY BINDING TO DNA.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; PROS-L (SHOWN HERE) AND PROS-
CC       S; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY HAVE DIFFERENT N-
CC       TERMINAL AMINO ACIDS OF THE HOMEODOMAIN.
CC   -!- TISSUE SPECIFICITY: NEURONAL PRECURSORS. EXPRESSED IN THE
CC       DEVELOPING CNS, LENS-SECRETING CONE CELLS OF THE EYE, AND MIDGUT.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN NEURONAL PRECURSORS EARLY DURING
CC       FORMATION.
CC   -!- SIMILARITY: BELONGS TO THE PROSPERO FAMILY OF HOMEOBOX PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; M81389; AAA28841.1; -.
DR   EMBL; D10609; BAA01464.1; -.
DR   EMBL; Z11743; CAA77802.1; -.
DR   PIR; A41089; A41089.
DR   PIR; JQ1397; JQ1397.
DR   FLYBASE; FBgn0004595; pros.
KW   Nuclear protein; Transcription regulation; DNA-binding; Homeobox;
KW   Developmental protein; Alternative splicing.
FT   DOMAIN        4     12       POLY-ALA.
FT   DOMAIN       28     31       POLY-SER.
FT   DOMAIN       32     35       POLY-ASN.
FT   DOMAIN      188    191       POLY-ALA.
FT   DOMAIN      253    260       POLY-GLN.
FT   DOMAIN      270    276       POLY-ASN.
FT   DOMAIN      282    286       POLY-ASN.
FT   DOMAIN      431    437       POLY-ASP.
FT   DOMAIN      505    508       POLY-ALA.
FT   DOMAIN      717    737       POLY-GLN.
FT   DOMAIN      754    763       POLY-GLN.
FT   DOMAIN      766    772       POLY-GLN.
FT   DOMAIN      934    937       POLY-ALA.
FT   DOMAIN      952    957       POLY-GLN.
FT   DOMAIN      960    963       POLY-GLN.
FT   DOMAIN      966    970       POLY-GLN.
FT   DOMAIN      991    998       NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN     1029   1048       POLY-GLN.
FT   DOMAIN     1074   1082       POLY-PRO.
FT   DNA_BIND   1241   1303       HOMEOBOX (ATYPICAL).
FT   DOMAIN     1304   1403       PROSPERO-LIKE.
FT   VARSPLIC   1216   1244       MISSING (IN ISOFORM PROS-S).
FT   CONFLICT     76     98       AKMLNELFGRQMKQAQDATSGLP -> GQDAERAVWPPDEA
FT                                GPGRNEWPA (IN REF. 1).
FT   CONFLICT    120    144       IGSLNSTSKLLQQQHNNNSIAPANS -> NLALQFHVQVAA
FT                                AAAITTALLPPIG (IN REF. 1).
FT   CONFLICT    418    418       H -> Q (IN REF. 2).
FT   CONFLICT    677    677       A -> C (IN REF. 1).
FT   CONFLICT    802    802       A -> S (IN REF. 1).
FT   CONFLICT    958    958       T -> S (IN REF. 1).
FT   CONFLICT   1048   1048       Q -> QQQQQ (IN REF. 1).
SQ   SEQUENCE   1403 AA;  153569 MW;  9EFB9973E24E238E CRC64;


  Query Match             36.9%;  Score 83;  DB 1;  Length 1403;
  Best Local Similarity   54.8%;  Pred. No. 0.18;
  Matches   17;  Conservative    7;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      10 ECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
           | ||: :||:::|| | ||   :|||||||:
Db     716 EQQQQQQQQQQQQQQQQQQEQQRRFEQEQQE 746


RESULT   6
TF2D_HUMAN
ID   TF2D_HUMAN     STANDARD;      PRT;   339 AA.
AC   P20226;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   TRANSCRIPTION INITIATION FACTOR TFIID (TATA-BOX FACTOR) (TATA
DE   SEQUENCE-BINDING PROTEIN) (TBP).
GN   TBP OR TFIID OR TF2D.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90302010; PubMed=2194289;
RA   Kao C.C., Lieberman P.M., Schmidt M.C., Zhou Q., Pei R., Berk A.J.;
RT   "Cloning of a transcriptionally active human TATA binding factor.";
RL   Science 248:1646-1649(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90326195; PubMed=2374612;
RA   Hoffmann A., Sinn E., Yamamoto T., Wang J., Roy A., Horikoshi M.,
RA   Roeder R.G.;
RT   "Highly conserved core domain and unique N terminus with presumptive
RT   regulatory motifs in a human TATA factor (TFIID).";
RL   Nature 346:387-390(1990).
RN   [3]
RP   DOMAINS.
RX   MEDLINE=90302006; PubMed=2363050;
RA   Peterson M.G., Tanese N., Pugh B.F., Tjian R.;
RT   "Functional domains and upstream activation properties of cloned
RT   human TATA binding protein.";
RL   Science 248:1625-1630(1990).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 159-337.
RX   MEDLINE=96209823; PubMed=8643494;
RA   Nikolov D.B., Chen H., Halay E.D., Hoffmann A., Roeder R.G.,
RA   Burley S.K.;
RT   "Crystal structure of a human TATA box-binding protein/TATA element
RT   complex.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:4862-4867(1996).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 159-339.
RX   MEDLINE=96346176; PubMed=8757291;
RA   Juo Z.S., Chiu T.K., Leiberman P.M., Baikalov I., Berk A.J.,
RA   Dickerson R.E.;
RT   "How proteins recognize the TATA box.";
RL   J. Mol. Biol. 261:239-254(1996).
CC   -!- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION
CC       OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II. TFIID BINDS
CC       SPECIFICALLY TO THE TATA BOX PROMOTER ELEMENT WHICH LIES CLOSE TO
CC       THE POSITION OF TRANSCRIPTION INITIATION.
CC   -!- SUBUNIT: BINDS DNA AS A MONOMER.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: THE C-TERMINAL 180 RESIDUES ARE EXTREMELY WELL
CC       CONSERVED IN ALL EUKARYOTIC TFIID.
CC   -!- SIMILARITY: WEAK, WITH BACTERIAL POLYMERASE SIGMA-FACTORS.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; M55654; AAA36731.1; -.
DR   EMBL; X54993; CAA38736.1; -.
DR   PIR; A34830; A34830.
DR   PIR; A34831; A34831.
DR   PDB; 1TGH; 01-AUG-96.
DR   PDB; 1CDW; 23-DEC-96.
DR   TRANSFAC; T00794; -.
DR   MIM; 600075; -.
DR   INTERPRO; IPR000814; -.
DR   PFAM; PF00352; TBP; 2.
DR   PRINTS; PR00686; TIFACTORIID.
DR   PROSITE; PS00351; TFIID; 2.
KW   Transcription regulation; DNA-binding; Nuclear protein; Repeat;
KW   3D-structure.
FT   DOMAIN       55     95       POLY-GLN.
FT   REPEAT      165    241       1.
FT   REPEAT      255    332       2.
FT   CONFLICT     58     61       MISSING (IN REF. 2).
SQ   SEQUENCE   339 AA;  37698 MW;  A61A578D972B970B CRC64;


  Query Match             36.4%;  Score 82;  DB 1;  Length 339;
  Best Local Similarity   41.0%;  Pred. No. 0.068;
  Matches   16;  Conservative   14;  Mismatches    9;  Indels    0;  Gaps    0;
```

```
Qy       2 EDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
           |: ||: :: ||: :||:::|| | ||:  :: :|:|||
Db      53 EEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 91


RESULT   7
APA4_PAPAN
ID   APA4_PAPAN     STANDARD;      PRT;   401 AA.
AC   Q28758;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   APOLIPOPROTEIN A-IV PRECURSOR (APO-AIV) (FRAGMENT).
GN   APOA4.
OS   Papio anubis (Olive baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RX   MEDLINE=93340170; PubMed=8101842;
RA   Hixson J.E., Kammerer C.M., Mott G.E., Britten M.L., Birnbaum S.,
RA   Powers P.K., Vandeberg J.L.;
RT   "Baboon apolipoprotein A-IV. Identification of Lys-76-->Glu that
RT   distinguishes two common isoforms and detection of length
RT   polymorphisms at the carboxyl terminus.";
RL   J. Biol. Chem. 268:15667-15673(1993).
CC   -!- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND
CC       CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN
CC       LIPASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR
CC       COMPONENT OF HDL AND CHYLOMICRONS.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED PRIMARILY IN THE INTESTINE.
CC       SECRETED IN PLASMA.
CC   -!- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH
CC       22-MER IS ACTUALLY A TANDEM ARRAY OF TWO, A AND B, RELATED 11-
CC       MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-
CC       HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY
CC       THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL
CC       ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.
CC   -!- POLYMORPHISM: THERE ARE TWO COMMON APOA-IV ISOFORMS, I (SHOWN
CC       HERE) AND E. THE I ISOFORM IS ASSOCIATED WITH HIGHER LEVELS OF
CC       HIGH DENSITY LIPOPROTEIN-C ON A HIGH CHOLESTEROL, SATURATED FAT
CC       DIET.
CC   -!- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L13174; AAA35379.1; -.
DR   HSSP; P02649; 1NFO.
DR   INTERPRO; IPR000074; -.
DR   PFAM; PF01442; Apolipoprotein; 1.
KW   Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal;
KW   Polymorphism.
FT   NON_TER       1      1
FT   SIGNAL       <1      4       POTENTIAL.
FT   CHAIN         5    401       APOLIPOPROTEIN A-IV.
FT   DOMAIN       17    314       13 X 22 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT       17     38       1.
FT   REPEAT       44     65       2.
FT   REPEAT       66     87       3.
FT   REPEAT       99    120       4.
FT   REPEAT      121    142       5.
FT   REPEAT      143    164       6.
FT   REPEAT      165    186       7.
FT   REPEAT      187    208       8.
FT   REPEAT      209    230       9.
FT   REPEAT      231    252       10.
FT   REPEAT      253    270       11.
FT   REPEAT      271    292       12.
FT   REPEAT      293    314       13.
FT   DOMAIN      356    394       GLU/GLN-RICH.
FT   VARIANT      80     80       K -> E (IN ISOFORM E).
SQ   SEQUENCE   401 AA;  46538 MW;  0A76D1284AA9837F CRC64;


  Query Match             36.4%;  Score 82;  DB 1;  Length 401;
  Best Local Similarity   47.5%;  Pred. No. 0.079;
  Matches   19;  Conservative    6;  Mismatches    9;  Indels    6;  Gaps    1;

Qy       1 PEDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
           ||  | : :| :|:  ||||:|| : ||      ||||||
Db     357 PEQQQEQQQEQEQQQEQQEEQQQQEQQQ------EQEQQQ 390


RESULT   8
KAPC_DICDI
ID   KAPC_DICDI     STANDARD;      PRT;   648 AA.
AC   P34099;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT (EC 2.7.1.37).
GN   PKAC OR PK2 OR PK3.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Dictyostelida; Dictyostelium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91323730; PubMed=1864510;
RA   Buerki E., Anjard C., Scholder J.-C., Reymond C.D.;
RT   "Isolation of two genes encoding putative protein kinases regulated
RT   during Dictyostelium discoideum development.";
RL   Gene 102:57-65(1991).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=93385090; PubMed=8373760;
RA   Anjard C., Etchebehere L., Pinaud S., Veron M., Reymond C.D.;
RT   "An unusual catalytic subunit for the cAMP-dependent protein kinase
RT   of Dictyostelium discoideum.";
RL   Biochemistry 32:9532-9538(1993).
RN   [3]
RP   CHARACTERIZATION.
RC   STRAIN=AX3;
RX   MEDLINE=93066311; PubMed=1332055;
RA   Mann S.K.O., Yonemoto W.M., Taylor S.S., Firtel R.A.;
RT   "DdPK3, which plays essential roles during Dictyostelium development,
RT   encodes the catalytic subunit of cAMP-dependent protein kinase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:10701-10705(1992).
CC   -!- FUNCTION: ESSENTIAL FOR DIFFERENTIATION AND FRUIT MORPHOGENESIS.
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN.
CC   -!- SUBUNIT: IN DICTYOSTELIUM THE HOLOENZYME IS A DIMER COMPOSED OF
CC       A REGULATORY (R) AND A CATALYTIC (C) SUBUNIT. IN THE PRESENCE OF
CC       CAMP IT DISSOCIATES INTO THE ACTIVE C SUBUNIT AND AN R MONOMER.
CC   -!- DEVELOPMENTAL STAGE: CAPK ACTIVITY IS LOW IN VEGETATIVELY GROWING
CC       AMOEBAE, INCREASES DURING DEVELOPMENT OF AGGREGATION AND REACHES
CC       A MAXIMUM AT CULMINATION.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       CAMP SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M38703; -; NOT_ANNOTATED_CDS.
DR   PIR; JQ1150; JQ1150.
```

```
DR   HSSP; P05132; 2CPK.
DR   DICTYDB; DD02030; PKAC.
DR   INTERPRO; IPR000719; -.
DR   INTERPRO; IPR000961; -.
DR   INTERPRO; IPR002290; -.
DR   PFAM; PF00069; pkinase; 1.
DR   PFAM; PF00433; pkinase_C; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; cAMP;
KW   Phosphorylation.
FT   DOMAIN       58     64       ASN-RICH.
FT   DOMAIN      136    223       GLN-RICH.
FT   DOMAIN      233    250       THR-RICH.
FT   DOMAIN      336    590       PROTEIN KINASE.
FT   NP_BIND     342    350       ATP (BY SIMILARITY).
FT   BINDING     365    365       ATP (BY SIMILARITY).
FT   ACT_SITE    459    459       BY SIMILARITY.
FT   MOD_RES     490    490       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   648 AA;  74458 MW;  D0F9B3A48C58D084 CRC64;


  Query Match             34.7%;  Score 78;  DB 1;  Length 648;
  Best Local Similarity   37.5%;  Pred. No. 0.3;
  Matches   15;  Conservative   12;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       1 PEDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
           |:  | : :: ||:  ||:::|||| ||:  :: :|  ||
Db     143 PQQQQPQQQQPQQQQPQQQQQQQPQQQQQPQQQLQQNNQQ 182


RESULT   9
Y192_HUMAN
ID   Y192_HUMAN     STANDARD;      PRT;  2124 AA.
AC   Q93074;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL PROTEIN KIAA0192 (FRAGMENT).
GN   KIAA0192.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RX   MEDLINE=96281124; PubMed=8724849;
RA   Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. V.
RT   The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 3:17-24(1996).
CC   -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D83783; BAA12112.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
FT   DOMAIN      599    602       POLY-SER.
FT   DOMAIN     1201   1207       POLY-GLY.
FT   DOMAIN     1998   2124       GLN-RICH.
FT   DOMAIN     1998   2023       POLY-GLN.
FT   DOMAIN     2028   2033       POLY-GLN.
FT   DOMAIN     2037   2070       POLY-GLN.
FT   DOMAIN     2090   2097       POLY-GLN.
SQ   SEQUENCE   2124 AA;  237207 MW;  255FB9419EC39F42 CRC64;


  Query Match             34.4%;  Score 77.5;  DB 1;  Length 2124;
  Best Local Similarity   34.7%;  Pred. No. 0.92;
  Matches   17;  Conservative   14;  Mismatches    9;  Indels    9;  Gaps    1;

Qy       1 PEDPQRRYEECQQECRQQEERQQ---------PQCQQRCLKRFEQEQQQ 40
             ||  |:: :: ||: :||:::||          | ||: |:: :|:|||
Db    1996 PEQQQQQQQQQQQQQQQQQQQQQQQQQQQYHIRQQQQQQILRQQQQQQQQ 2044


RESULT  10
YM38_YEAST
ID   YM38_YEAST     STANDARD;      PRT;   758 AA.
AC   Q03825;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 85.0 KDA PROTEIN IN HLJ1-SMP2 INTERGENIC REGION.
GN   YMR164C OR YM8520.13C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Hunt S., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z49705; CAA89800.1; -.
DR   SGD; S0004774; YMR164C.
KW   Hypothetical protein.
FT   DOMAIN      290    329       POLY-GLN.
FT   DOMAIN      605    637       POLY-ASN.
FT   DOMAIN      653    656       POLY-SER.
SQ   SEQUENCE   758 AA;  85050 MW;  BA05BFC754D9294B CRC64;


  Query Match             34.2%;  Score 77;  DB 1;  Length 758;
  Best Local Similarity   40.5%;  Pred. No. 0.43;
  Matches   15;  Conservative   13;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       4 PQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
             ||:: :: ||: :||:::|| | ||:  ::  |:|||
Db     293 PQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHQQQQQ 329


RESULT  11
INVO_MOUSE
ID   INVO_MOUSE     STANDARD;      PRT;   467 AA.
AC   P48997;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   INVOLUCRIN.
GN   IVL.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIH SWISS;
RX   MEDLINE=94104476; PubMed=8277848;
RA   Djian P., Phillips M., Easley K., Huang E., Simon M., Rice R.H.,
```

```
RA   Green H.;
RT   "The involucrin genes of the mouse and the rat: study of their shared
RT   repeats.";
RL   Mol. Biol. Evol. 10:1136-1149(1993).
CC   -!- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC       IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC       MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC       FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC   -!- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC       OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L28819; AAA39330.1; -.
DR   MGD; MGI:96626; IVL.
DR   INTERPRO; IPR002360; -.
DR   PROSITE; PS00795; INVOLUCRIN; 1.
KW   Keratinocyte; Repeat.
SQ   SEQUENCE   467 AA;  54919 MW;  603E1E51B435737D CRC64;


  Query Match             34.0%;  Score 76.5;  DB 1;  Length 467;
  Best Local Similarity   33.3%;  Pred. No. 0.32;
  Matches   16;  Conservative   14;  Mismatches    9;  Indels    9;  Gaps    1;

Qy       2 EDPQRRYEECQQECRQQEERQQPQCQQR---------CLKRFEQEQQQ 40
           ::|| : :  :|: :||:::|| | ||:         || : :|:||| 
Db     123 QEPQMQEQHLRQQQQQQQQQQQQQQQQQQQQETQEQGLCLGQKQQQQQQ 170


RESULT  12
YKF4_YEAST
ID   YKF4_YEAST     STANDARD;      PRT;   738 AA.
AC   P35732;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   HYPOTHETICAL 84.0 KDA PROTEIN IN NUP120-CSE4 INTERGENIC REGION.
GN   YKL054C OR YKL308.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=94378723; PubMed=8091862;
RA   Rasmussen S.W.;
RT   "Sequence of a 28.6 kb region of yeast chromosome XI includes the
RT   FBA1 and TOA2 genes, an open reading frame (ORF) similar to a
RT   translationally controlled tumour protein, one ORF containing motifs
RT   also found in plant storage proteins and 13 ORFs with weak or no
RT   homology to known proteins.";
RL   Yeast 10:S63-S68(1994).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X75781; CAA53418.1; -.
DR   EMBL; Z28054; CAA81890.1; -.
DR   PIR; S37876; S37876.
DR   PIR; S39184; S39184.
DR   SGD; S0001537; YKL054C.
KW   Hypothetical protein.
FT   DOMAIN      382    718       GLN-RICH.
FT   DOMAIN      574    580       POLY-ALA.
FT   DOMAIN      708    716       POLY-ALA.
SQ   SEQUENCE   738 AA;  83973 MW;  05734E2D0B7389AC CRC64;


  Query Match             34.0%;  Score 76.5;  DB 1;  Length 738;
  Best Local Similarity   41.5%;  Pred. No. 0.47;
  Matches   17;  Conservative   12;  Mismatches   11;  Indels    1;  Gaps    1;

Qy       1 PEDPQRRYEECQ-QECRQQEERQQPQCQQRCLKRFEQEQQQ 40
             |: ||:  :  | |: :||:: |||| |: |:: :|:||| 
Db     390 PQQPQQPQQPQQPQQPQQQQQPQQPQQPQQQLQQQQQQQQQ 430


RESULT  13
SNF5_YEAST
ID   SNF5_YEAST     STANDARD;      PRT;   905 AA.
AC   P18480;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   TRANSCRIPTION REGULATORY PROTEIN SNF5 (SWI/SNF COMPLEX COMPONENT SNF5)
DE   (TRANSCRIPTION FACTOR TYE4).
GN   SNF5 OR TYE4 OR SWI10 OR YBR289W OR YBR2036.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MCY;
RX   MEDLINE=91042489; PubMed=2233708;
RA   Laurent B.C., Treitel M.A., Carlson M.;
RT   "The SNF5 protein of Saccharomyces cerevisiae is a glutamine- and
RT   proline-rich transcriptional activator that affects expression of a
RT   broad spectrum of genes.";
RL   Mol. Cell. Biol. 10:5616-5625(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=94378722; PubMed=8091861;
RA   Holmstroem K., Brandt T., Kallesoe T.;
RT   "The sequence of a 32,420 bp segment located on the right arm of
RT   chromosome II from Saccharomyces cerevisiae.";
RL   Yeast 10:S47-S62(1994).
CC   -!- FUNCTION: INVOLVED IN TRANSCRIPTIONAL ACTIVATION. THE SWI/SNF
CC       COMPLEX IS REQUIRED FOR THE INDUCED EXPRESSION OF A LARGE NUMBER
CC       OF GENES. THIS COMPLEX ALTERS CHROMATIN STRUCTURE TO FACILITATE
CC       BINDING OF GENE-SPECIFIC DEDICATED TRANSCRIPTION FACTORS.
CC   -!- SUBUNIT: COMPONENT OF THE SWI/SNF GLOBAL TRANSCRIPTION ACTIVATOR
CC       COMPLEX.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE SNF5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M36482; AAA35062.1; -.
DR   EMBL; X76053; CAA53652.1; -.
DR   EMBL; Z36158; CAA85254.1; -.
DR   PIR; S44551; RGBYS5.
DR   PIR; S39145; S39145.
DR   SGD; S0000493; SNF5.
KW   Transcription regulation; Activator; Nuclear protein.
FT   DOMAIN       31    270       GLN-RICH.
FT   DOMAIN       72    132       PRO-RICH.
FT   DOMAIN      272    324       PRO-RICH.
```

```
FT   DOMAIN      489    588       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      714    882       PRO-RICH.
FT   DOMAIN      755    798       ARG/LYS-RICH (BASIC).
FT   CONFLICT    564    564       E -> D (IN REF. 1).
SQ   SEQUENCE   905 AA;  102557 MW;  A287B4A648DD1A35 CRC64;


  Query Match             33.8%;  Score 76;  DB 1;  Length 905;
  Best Local Similarity   35.9%;  Pred. No. 0.63;
  Matches   14;  Conservative   16;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      2 EDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
          :  |::::: ||: :||:::|| | ||:  :: :|:|||
Db    225 QQQQQQHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 263


RESULT  14
CLOC_DROME
ID   CLOC_DROME     STANDARD;      PRT;  1023 AA.
AC   O61735; O76342; O77137;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CIRCADIAN LOCOMOTER OUTPUT CYCLES KAPUT PROTEIN (DCLOCK) (DPAS1).
GN   CLK OR JRK OR CLOCK OR PAS1.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEAD;
RX   MEDLINE=98279147; PubMed=9616122;
RA   Darlington T.K., Wager-Smith K., Ceriani M.F., Staknis D., Gekakis N.,
RA   Steeves T.D.L., Weitz C.J., Takahashi J.S., Kay S.A.;
RT   "Closing the circadian loop: CLOCK-induced transcription of its own
RT   inhibitors per and tim.";
RL   Science 280:1599-1603(1998).
RN   [2]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS.
RC   TISSUE=HEAD;
RX   MEDLINE=98292177; PubMed=9630223;
RA   Allada R., White N.E., So W.V., Hall J.C., Rosbash M.;
RT   "A mutant Drosophila homolog of mammalian Clock disrupts circadian
RT   rhythms and transcription of period and timeless.";
RL   Cell 93:791-804(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=98414630; PubMed=9742131;
RA   Bae K., Lee C., Sidote D., Chuang K.-Y., Edery I.;
RT   "Circadian regulation of a Drosophila homolog of the mammalian clock
RT   gene: PER and TIM function as positive regulators.";
RL   Mol. Cell. Biol. 18:6142-6151(1998).
CC   -!- FUNCTION: CIRCADIAN REGULATOR THAT ACTS AS A TRANSCRIPTION FACTOR
CC       AND GENERATES A RYTHMIC OUTPUT WITH A PERIOD OF ABOUT 24 HOURS.
CC       OSCILLATES IN ANTIPHASE TO THE CYCLING OBSERVED FOR PERIOD (PER)
CC       AND TIMELESS (TIM). ACCORDING TO REF.3, REACHES PEAK ABUNDANCE
CC       WITHIN SEVERAL HOURS OF THE DARK-LIGHT TRANSITION AT ZT0
CC       (ZEITGEBER 0), WHEREAS REF.1 DESCRIBES BIMODAL OSCILLATING
CC       EXPRESSION WITH MAXIMUM AT ZT5 AND ZT23. CLOCK-CYCLE HETERODIMERS
CC       ACTIVATE CYCLING TRANSCRIPTION OF PER AND TIM BY BINDING TO THE E-
CC       BOX (3'-CACGTG-5') PRESENT IN THEIR PROMOTERS. ONCE INDUCED,
CC       PERIOD AND TIMELESS BLOCK CLOCK'S ABILITY TO TRANSACTIVATE THEIR
CC       PROMOTERS.
CC   -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC       BHLH PROTEIN. FORMS A HETERODIMER WITH CYCLE.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; THE FULL-LENGTH VARIANT A
CC       (SHOWN HERE) AND VARIANT B; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC       VARIANT B ENCODES TWO CONCEPTUAL PROTEINS, THE FIRST CONSISTS ONLY
CC       OF THE BHLH DOMAIN, THE OTHER CONSISTS OF THE PAS-1 AND ALL C-
CC       TEMINAL DOMAINS. VARIANT B IS EXPRESSED WEAKLY AT ALL THE TIME OF
CC       THE DAY, AND IT CYCLES IN PHASE WITH THE FULL-LENGTH FORM.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. FOUND IN HEAD, BODY, AND
CC       APPENDAGE FRACTIONS.
CC   -!- DOMAIN: CONTAINS THREE POLYGLUTAMINE REPEATS WHICH COULD
CC       CORRESPOND TO THE TRANSACTIVATION DOMAIN. THE LENGHT OF THE
CC       REPEATS IS POLYMORPHIC. IN THE ARRYTHMIC MUTANT JRK, DELETION OF
CC       THIS REGION LEADS TO THE LOSS OF CIRCADIAN RHYTHMICITY AND ALTERED
CC       LIGHT RESPONSE.
CC   -!- POLYMORPHISM: THE VARIABILITY IN LENGTH OF THE POLYGLUTAMINE
CC       STRETCH IS DUE TO POLYMORPHISM OF THIS REGION.
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC       TRANSCRIPTION FACTORS.
CC   -!- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF067207; AAD10630.1; -.
DR   EMBL; AF065133; AAC39101.1; -.
DR   EMBL; AF069997; AAC62234.1; -.
DR   FLYBASE; FBgn0023076; clk.
DR   INTERPRO; IPR000014; -.
DR   INTERPRO; IPR001067; -.
DR   INTERPRO; IPR001092; -.
DR   INTERPRO; IPR003015; -.
DR   PFAM; PF00010; HLH; 1.
DR   PFAM; PF00989; PAS; 2.
DR   PRINTS; PR00785; NCTRNSLOCATR.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW   Transcription regulation; Nuclear protein; Repeat; Biological rhythms;
KW   DNA-binding; Alternative splicing.
FT   DNA_BIND     12     24       BASIC DOMAIN.
FT   DOMAIN       25     62       HELIX-LOOP-HELIX MOTIF.
FT   REPEAT       86    152       PAS-1.
FT   REPEAT      251    317       PAS-2.
FT   DOMAIN      548    559       POLY-GLN.
FT   DOMAIN      766    769       POLY-GLN.
FT   DOMAIN      794    836       POLY-GLN.
FT   DOMAIN      874    877       POLY-ASN.
FT   DOMAIN      887    895       POLY-ASN.
FT   DOMAIN      953    963       POLY-GLN.
FT   DOMAIN      776   1023       IMPLICATED IN THE CIRCADIAN RHYTHMICITY.
FT   VARIANT     816    823       MISSING.
FT   CONFLICT     12     12       K -> KSFLC (IN REF. 3).
FT   CONFLICT     32     32       N -> D (IN REF. 3).
FT   CONFLICT    128    128       N -> K (IN REF. 2).
FT   CONFLICT    555    555       N -> S (IN REF. 1).
FT   CONFLICT    605    605       I -> L (IN REF. 3).
FT   CONFLICT    912    912       Y -> C (IN REF. 3).
SQ   SEQUENCE   1023 AA;  115751 MW;  514374CBC050DAFB CRC64;


  Query Match             33.3%;  Score 75;  DB 1;  Length 1023;
  Best Local Similarity   38.9%;  Pred. No. 0.88;
  Matches   14;  Conservative   15;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      5 QRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
            |::::: ||: :||:::|| | ||:  :: :|:|||
Db    794 QQQHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 829


RESULT  15
WC1_NEUCR
ID   WC1_NEUCR      STANDARD;      PRT;  1154 AA.
AC   Q01371;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
```

```
DE   WHITE COLLAR 1 PROTEIN (WC1).
GN   WC-1.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=74-OR23-1A;
RX   MEDLINE=96203083; PubMed=8612589;
RA   Ballario P., Vittorioso P., Magrelli A., Talora C., Cabibbo A.,
RA   Macino G.;
RT   "White collar-1, a central regulator of blue light responses in
RT   Neurospora, is a zinc finger protein.";
RL   EMBO J. 15:1650-1657(1996).
CC   -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR INVOLVED IN LIGHT
CC       REGULATION. BINDS AND AFFECTS BLUE LIGHT REGULATION OF THE AL-3
CC       GENE. WC1 AND WC2 PROTEINS INTERACT VIA HOMOLOGOUS PAS DOMAINS,
CC       BIND TO PROMOTERS OF LIGHT REGULATED GENES SUCH AS FRQ, AND
CC       ACTIVATE TRANSCRIPTION.
CC   -!- SUBUNIT: HETERODIMER OF WC1 AND WC2 (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- INDUCTION: BY BLUE LIGHT.
CC   -!- DOMAIN: THE GLUTAMINE-RICH DOMAIN MIGHT FUNCTION IN ACTIVATING
CC       GENE EXPRESSION.
CC   -!- SIMILARITY: BELONGS TO THE GATA-TYPE TRANSCRIPTION FACTOR FAMILY.
CC   -!- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X94300; CAA63964.1; -.
DR   HSSP; P17678; 1GAU.
DR   INTERPRO; IPR000014; -.
DR   INTERPRO; IPR000679; -.
DR   INTERPRO; IPR001610; -.
DR   PFAM; PF00320; GATA; 1.
DR   PFAM; PF00785; PAC; 1.
DR   PFAM; PF00989; PAS; 3.
DR   PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR   PROSITE; PS50114; GATA_ZN_FINGER_2; 1.
KW   Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW   Nuclear protein; Repeat.
FT   DOMAIN       16     61       GLN-RICH.
FT   REPEAT      381    448       PAS-1.
FT   DOMAIN      469    508       PAC MOTIF.
FT   REPEAT      576    642       PAS-2.
FT   DOMAIN      650    691       PAC MOTIF.
FT   REPEAT      695    761       PAS-3.
FT   ZN_FING     935    960       GATA-TYPE.
FT   DOMAIN       21     57       POLY-GLN.
FT   DOMAIN      329    333       POLY-PRO.
SQ   SEQUENCE   1154 AA;  125944 MW;  CB4CB26047ECCDC5 CRC64;


  Query Match             33.3%;  Score 75;  DB 1;  Length 1154;
  Best Local Similarity   38.5%;  Pred. No. 0.98;
  Matches   15;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 PEDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQ 39
             ||: | :  : ||: :||:::|| | ||:  :: :|:||
Db      12 PEELQHQMHQHQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 50


Search completed: March  1, 2001, 16:16:57
Job time: 437 sec
```